DATA-DRIVEN BACKGROUND ESTIMATION FOR THE SUSY SEARCH AT ATLAS

Y. KATAOKA (FOR THE ATLAS COLLABORATION)
International Center for Elementary Particle Physics, the University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
E-mail: Yousuke.Kataoka@cern.ch

The Large Hadron Collider (LHC) will start operation in 2008 at CERN. One of the major topics in LHC is supersymmetry (SUSY) search. The key point of the SUSY search is the understanding of the Standard Model (SM) backgrounds. Since the background estimation based on Monte Carlo has large uncertainty and one can also expect the uncertainties from detector performance especially at early stage, we have been developing the background estimation techniques using real data[5]. We describe the data-driven techniques of background estimation for SUSY search at ATLAS.

1. Introduction

Supersymmetry (SUSY) is one of the most attractive extension of the Standard Model (SM) [1–3]. It solves the hierarchy problem and brings a chance of gauge unification at higher energy. Moreover, the lightest SUSY particle is a good candidate of the cold dark matter. The Large Hadron Collider (LHC) is considered to be a crucial experiment for the SUSY search due to its high center of mass energy (14 TeV) and high luminosity, and ATLAS has the discovery potential of 1TeV SUSY at early data (1 fb^{-1})[4,5].

2. Inclusive search

The cross sections of the strongly interacting SUSY particles such as gluinos and squarks are large at the LHC, if these particles are not too massive. If R-parity is conserved, these particles decay into the lightest SUSY particle via cascade decays. The lightest SUSY particles escape the detector unseen and give rise to large missing E_T, and other products are also detected as jets or leptons. Thus large missing E_T and multi jets are fairly model independent signatures of SUSY events and our inclusive search only relies on them.

We require the following baseline cuts for the inclusive search[4].

- 4 jets ($p_T > 50$ GeV)
- 1 jet ($p_T > 100$ GeV)
- missing $E_T > 100$ GeV
- missing $E_T > 0.2\times$ M$_{eff}$
- Transverse Sphericity > 0.2

The Effective Mass (M$_{eff}$) above is defined by the following equation.

$$\mathrm{M}_{eff} = \sum_{\text{4jets}} pT \text{ of jet} + \text{missing } E_T \tag{1}$$

Since M$_{eff}$ includes both multi jets and missing E_T signatures, it is used as a discriminating variable for the inclusive search.

For no lepton mode, we require the additional $\Delta\phi$ cut below

- $\Delta\phi_{123}$ (minimum $\Delta\phi$ between the missing E_T vector and the direction of the three leading jets)< 0.2

The $\Delta\phi_{123}$ distribution for the no lepton mode is shown in Figure 1. The $\Delta\phi$ cut highly suppresses QCD background since missing E_T come from semileptonic decay of heavy flavor quark or mis-measurement of jet energy and is mostly correlated with the direction of the jet.

For one lepton mode, we require one isolated electron or muon with $p_T > 20$ GeV, which is included in the M$_{eff}$ calculation, and also require the following Transverse Mass (M$_T$) cut on top of the baseline cuts

- M$_T$ >100 GeV

The M$_T$ distribution for the one lepton mode is shown in Figure 1. The missing ET is constrained with the lepton p_T to have a transverse mass no larger than the W mass, and the M$_T$ cut eliminates most background, except for the di-leptonic $t\bar{t}$ background where both W's can decay leptonically.

Figure 2 shows M$_{eff}$ distributions for no and one lepton mode at 1 fb^{-1}. The no lepton mode is statistically preferable but data-driven background estimation is rather complicated since all W, Z, $t\bar{t}$ and QCD backgrounds contribute. On the other hand, the background of one lepton mode is mostly di-leptonic $t\bar{t}$ background and almost free from QCD background, which make the background estimation less complicated. In this sense, no and one lepton modes are complementary.

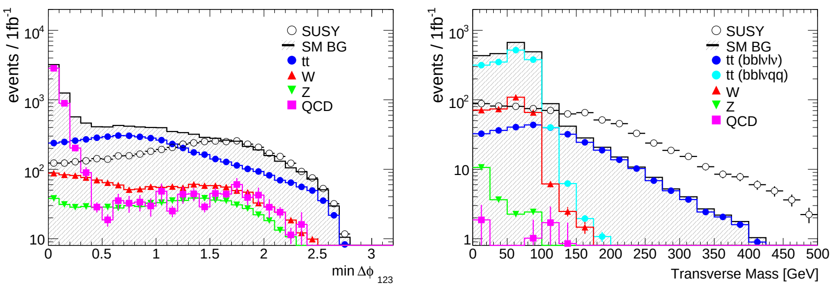

Figure 1. $\Delta\phi_{123}$ (minimum $\Delta\phi$ between the missing E_T vector and the direction of the three leading jets) distribution of no lepton mode (left figure) and M_T distribution of one lepton mode (right figure). The hatched histogram shows Standard Model background. $t\bar{t}$,W,Z and QCD background are also separately shown. The SUSY signal (squark and gluino mass ∼700 GeV) is superimposed as a reference.

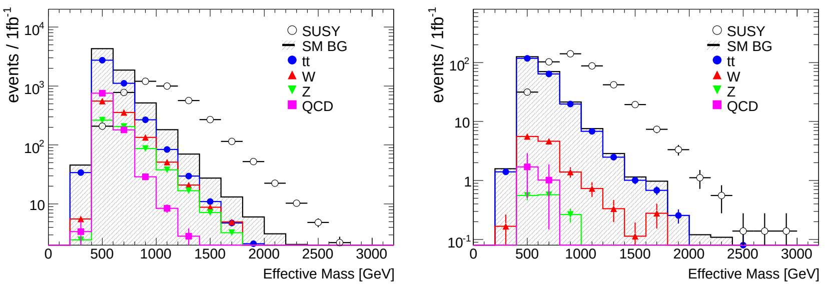

Figure 2. Effective Mass (M_{eff}) distributions for no lepton mode (left figure) and one lepton mode (right figure). The hatched histogram shows Standard Model background. $t\bar{t}$,W,Z and QCD background are also separately shown. The SUSY signal (mass scale∼700 GeV) is superimposed as a reference.

3. Data-driven background estimation

3.1. *$Z \to \nu\nu$ background estimation (replace method)*

$Z \to ll$ + multi-jets data is a good control sample for $Z \to \nu\nu$ background estimation. After the selection of the di-lepton control sample (two electrons or muons of opposite charge with $p_T > 20$ GeV, missing $E_T < 40$ GeV, M_{ll} within ± 10 GeV from M_Z, the contamination of $t\bar{t}$ background is negligibly small($\sim 1\%$). Then we replace ll with missing energy and apply the no lepton mode SUSY cuts, which ensures the same kinematic condition with the real background except for the di-lepton efficiency. With the lepton efficiency correction, the distribution must reproduce the real $Z \to \nu\nu$ background. However, the statistics of the control data (around 100 events at 1 fb^{-1}) is critical for this method. Since studies have shown that the uncertainties of the Monte Carlo predictions mainly affect the Monte Carlo normalization, and not the shape of the missing E_T and M_{eff} distributions, additional information can be gained by using Monte Carlo shape, and normalizing to data. This is why we still have most of the advantages of data-driven technique by this method.

3.2. *QCD background estimation ($\Delta\phi$ and smearing methods)*

The normalization of the QCD background can be estimated using the peak close to zero in the $\Delta\phi_{123}$ distribution. The contamination of Standard Model background, and possibly of SUSY events, can be subtracted using the side band ($0.4 < \Delta\phi_{123} < 0.6$). In principle, the missing E_T and M_{eff} shape of the QCD background can also be extracted from the peak events after a side band subtraction. However, the estimated normalization and shape using this procedure are those before the $\Delta\phi$ cut, and we would still have to rely on Monte Carlo to estimate the background after this cut. Therefore, to estimate the missing E_T and M_{eff} shape of the QCD background, we use another method. We note that the dominant source of background after the $\Delta\phi$ cut is heavy flavor jets, with an energetic neutrino from heavy quark decay. We select a multi-jet control sample (four jets with $p_T > 50$ GeV, missing $E_T < 50$ GeV), and smear the jets with a "missing energy fraction" extracted either from Monte Carlo, or from a sample of data events with significant, but non-isolated, missing E_T. After applying the SUSY selection cuts on the smeared events, we have the estimated background shape.

3.3. *W and $t\bar{t}$ background estimation (M_T method)*

Since W and $t\bar{t}$ backgrounds resemble each other in their features, we estimate W and $t\bar{t}$ backgrounds inclusively. In one lepton mode, M_T cut divides data into signal region and background dominant region. The latter events can be used to estimate the background shape since correlation between M_T and missing E_T (or M_{eff}) is small. Figure 3 shows the comparison between the true background in the one lepton mode, and the estimation using the control sample, in which the normalization is determined in the low missing E_T region between 100 GeV and 200 GeV.

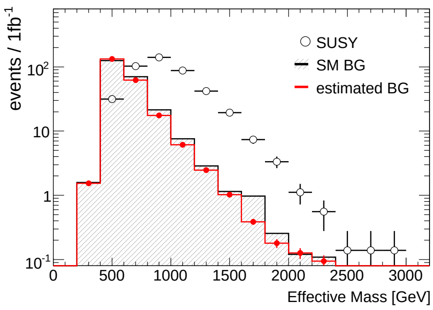

Figure 3. Background estimation by M_T method. The hatched histogram shows M_{eff} distribution of one lepton mode background. The estimated background (red line) agrees well with the real background.

The estimated background agrees well with the real background. As with other data-driven methods, the estimated background doesn't rely on Monte Carlo in both its shape and normalization. The other advantage is that the estimation is fairly stable against detector uncertainty such as jet energy scale. For no lepton mode and di-lepton mode, the same background shape can be used although for no lepton mode the contamination of Z and QCD backgrounds have to be subtracted at the normalization. This is because $t\bar{t}$ background is dominant through no, one, di-lepton mode and the background shape is mostly determined by the kinematics of $t\bar{t}$.

4. Conclusions

We have developed a number of data-driven background estimation techniques for the SUSY search at ATLAS. The several methods described above cover all the backgrounds of no, one, di-lepton mode. The estimations are less dependent on Monte Carlo and also stable against uncertainties from detector performance.

Acknowledgments

The author thanks the ATLAS collaboration especially SUSY Working Group for their support of this work. The special appreciation goes to the ATLAS group of University of Tokyo for their joint studies.

References

1. J. Wess and B. Zumino, *Nucl Phys.* **B70**, 39 (1974)

2. P. Fayet and S. Ferrara, *Phys. Rep.* **32**, 249 (1977)
3. H. P. Nilles, *Phys. Rep.* **110**, 1 (1984)
4. ATLAS collaboration, ATLAS Detector and Physics Performance Technical Design Report, CERN/LHCC/99-15 (1999)
5. ATLAS Collaboration, Expected performance of the ATLAS experiment - Detector, Trigger and Physics-, CERN report (2008), in preparation.